

Ray Hyman - The Evidence for Psychic Functioning: Claims vs. Reality

Skeptical Inquirer magazine : March/April 1996

<http://www.csicop.org/si/9603/claims.html>

The recent media frenzy over the Stargate report violated the truth. Sober scientific assessment has little hope of winning in the public forum when pitted against unsubstantiated and unchallenged claims of "psychics" and psychic researchers -- especially when the claimants shamelessly indulge in hyperbole. While this situation may be depressing, it is not unexpected. The proponents of the paranormal have seized an opportunity to achieve by propaganda what they have failed to achieved through science.

Most of these purveyors of psychic myths should not be taken seriously. However, when one of the persons making extreme claims is Jessica Utts, who is a professor of statistics at the University of California at Davis, this is another matter. Utts has impressive credentials and she marshals the evidence for her case in an effective way. So it is important to look at the basis for what I believe are extreme claims, even for a parapsychologist. Here is what Utts writes in her report on the Stargate program:

"Using the standards applied to any other area of science, it is concluded that psychic functioning has been well established. The statistical results of the studies examined are far beyond what is expected by chance. Arguments that these results could be due to methodological flaws in the experiments are soundly refuted. Effects of similar magnitude to those found in government-sponsored research at SRI [Stanford Research Institute] and SAIC [Science Applications International Corporation] have been replicated at a number of laboratories across the world. Such consistency cannot be readily explained by claims of flaws or fraud. . . . [Psychic functioning] is reliable enough to be replicated in properly conducted experiments, with sufficient trials to achieve the long-run statistical results needed for replicability. . . . Precognition, in which the answer is known to no one until a future time, appears to work quite well. . . . There is little benefit to continuing experiments designed to offer proof, since there is little more to be offered to anyone who does not accept the current collection of data."

For what it is worth, I happen to be one of those "who does not accept the current collection of data" as proving psychic functioning. Indeed, I do not believe that "the current collection of data" justifies that an anomaly of any sort has been demonstrated, let alone a paranormal anomaly. Although Utts and I -- in our capacities as coevaluators of the Stargate project -- evaluated the same set of data, we came to very different conclusions. If Utts's conclusion is correct, then the fundamental principles that have so successfully guided the progress of science from the days of Galileo and Newton to the present must be drastically revised. Neither relativity theory nor quantum mechanics in their present versions can cope with a world that harbors the psychic phenomena so boldly proclaimed by Utts and her parapsychological colleagues.

So, it is worth looking at the evidence that Utts uses to buttress her case. Unfortunately, many of the issues that this evidence raises are technical or require long and tedious refutations. This is not the place to develop this lengthy rebuttal. Instead, I will briefly list the sources of Utts's evidence and try to provide at least one or two simple reasons why they do not, either singly or taken together, justify her conclusions. As I understand it, Utts supports her conclusion with the following sources of evidence:

1. Meta-analyses of Previous Parapsychological Experiments

In a meta-analysis, an investigator uses statistical tools to pool the data from a series of similar experiments published over a period of time that may involve several different investigators and laboratories. Although some or many of the individual experiments might have yielded weak or nonsignificant results, the pooled data can be highly significant from a statistical viewpoint. In addition to getting an overall measure of significance, the meta-analyses typically also grade each study for quality on one or more dimensions. The idea is to see if the successful outcomes are correlated with poor quality. If so, this counts against the evidence for paranormal functioning. If not, then this is proclaimed as evidence that the successful outcomes were not due to flaws.

In the four major meta-analyses of previous parapsychological research, the pooled data sets produced astronomically significant results while the correlation between successful outcome and rated quality of the experiments was essentially zero.

Much can be written at this point. The major point I would make, however, is that drawing conclusions from meta-analytic studies is like having your cake and eating it too. The same data are being used to generate and test a hypothesis. The proper use of meta-analysis is to generate hypotheses, which then must be independently tested on new data. As far as I know, this has yet to be done. The correlation between quality and outcome also must be suspect because the ratings are not done blindly.

As far as I can tell, I was the first person to do a meta-analysis on parapsychological data. I did a meta-analysis of the original ganzfeld experiments as part of my critique of those experiments. My analysis demonstrated that certain flaws, especially quality of randomization, did correlate with outcome. Successful outcomes correlated with inadequate methodology. In his reply to my critique, Charles Honorton did his own meta-analysis of the same data. He too scored for flaws, but he devised scoring schemes different from mine. In his analysis, his quality ratings did not correlate with outcome. This came about because, in part, Honorton found more flaws in unsuccessful experiments than I did. On the other I found more flaws in successful experiments than Honorton did. Presumably, both Honorton and I believed we were rating quality in an objective and unbiased way. Yet, both of us ended up with results that matched our preconceptions.

So far, other than my meta-analysis, all the meta-analyses evaluating quality and outcome have been carried out by parapsychologists. We might reasonably expect that the findings will differ with skeptics as raters.

These are just two, but very crucial, reasons why the meta-analyses conducted so far on parapsychological data cannot be used as evidence for psi.

2. The Original Ganzfeld Experiments

These consisted of 42 experiments (by Honorton's count) of which 55 percent had been claimed as producing significant results in favor of ESP. My meta-analysis and evaluation of these experiments showed that this database did not justify concluding that ESP was demonstrated. Honorton's meta-analysis and rebuttal suggests otherwise. Utts naturally relies on Honorton's meta-analysis and ignores mine. In our joint paper, both Honorton and I agreed that there were sufficient problems with this original database that nothing could be concluded until further replications, conducted according to specified criteria, appeared.

3. The Autoganzfeld Experiments

This series of experiments, conducted over a period of six years, is so named because the collection of data was partially automated. When this set of experiments was first published in the *Journal of Parapsychology* in 1990, it was presented as a successful replication of the original ganzfeld experiments. Moreover, these experiments were said to have been conducted according to the criteria set out by Honorton and me. This indeed seemed to be the case with the strange exception of the procedure for randomizing targets at presentation and judging. Even in writing our joint paper, Honorton argued with me that careful randomization was not necessary in the ganzfeld experiments because each subject appears only once. I disagreed with Honorton, but even by his own reasoning, randomization is not as important if you believe that the subject is the sole source of the final judgment. But this was blatantly not the case in the autoganzfeld experiments. The experimenter, who was not so well shielded from the sender as the subject, interacted with the subject during the judging process. Indeed, during half of the trials the experimenter deliberately prompted the subject during the judging procedure. This means that the judgments from trial to trial were not strictly independent.

However, from the original published report, I had little reason to question the methodology of these experiments. What I did question was the claim that they were consistent with the original ganzfeld experiments. I pointed out a number of ways that the two outcomes were inconsistent. Not until I was asked to write a response to a new presentation of these experiments in the January 1994 issue of the *Psychological Bulletin* did I get an opportunity to scrutinize the raw data. Unfortunately, I did not get all of the data, especially the portion that I needed to make direct tests of the randomizing procedures. But my analyses of what I did get uncovered some peculiar and strong patterns in the data. All of the significant hitting was done on the second or later appearance of a target. If we examined the guesses against just the first occurrences of targets, the result is consistent with chance. Moreover, the hit rate rose systematically with each additional occurrence of a target. This suggests to me a possible flaw. Daryl Bem, the coauthor with Honorton of the *Psychological Bulletin* paper, responded that it might reveal another peculiarity of psychic phenomena. The reason why my finding is of concern is that all the targets were on videotape and played on tape players during presentation. At the very least, the peculiar pattern I identified suggests that we need to require that when targets and decoys are presented to the subjects for judging, they all have been run through the machine the exact same number of times. Otherwise there might be nonparanormal reasons why one of the video clips appears different to the subjects.

Subsequent to my response, I have learned about other possible problems with the autoganzfeld experiments. The point of this is to show that it takes time and critical scrutiny to realize that what at first seems like an airtight series of experiments has a variety of possible weaknesses. I concluded, and do so even more strongly now, that the autoganzfeld experiments constitute neither a successful replication of the original ganzfeld experiments nor a sufficient body of data to conclude that ESP has finally been demonstrated. This new set of experiments needs independent replication with tighter controls.

4. Apparent Replications of the Autoganzfeld Experiments

Utts points to some apparent replications of the ganzfeld experiments that have been reported at parapsychological meetings. The major one is a direct attempt to replicate the autoganzfeld experiments with better controls, done at the University of Edinburgh. The

reported results were apparently significant but were due to just one of the three experimenters. The two experienced experimenters produced only chance hitting. There are some inconsistencies in these unpublished reports. Utts points to three different replications that were apparently successful. I have heard of at least two large-scale replications that were unsuccessful. None of these replications, however, has been reported in a refereed journal and none has had the opportunity to be critically scrutinized. So we cannot count these one way or the other at this time until we know the details.

5. The SAIC Experiments

Utts and I were hired as the evaluation panel to assess the results of 20 years of previously classified research on remote viewing and related ESP phenomena. In the time available to us, it was impossible to scrutinize carefully all the documents generated by this program. Instead, we focused our efforts on evaluating the ten studies done at Science Applications International Corporation (SAIC) during the early 1990s. These were selected, in consultation with the principal investigator, as representing the best experiments in the set. These ten experiments included two that examined physiological correlates of ESP. The results were negative. Another study found a correlation between when a subject was being observed (via remote camera) and galvanic skin reactions. The remaining studies, in one way or another, dealt with various target and other factors that might influence remote viewing ability. In these studies the same set of viewers produced descriptions that were successfully matched against the correct target consistently better than chance (with some striking exceptions).

Neither Utts nor I had the time or resources to fully scrutinize the laboratory procedures or data from these experiments. Instead, we relied on what we could glean from reading the technical reports. Two of the experiments had recently been published in the *Journal of Parapsychology*. The difficulty here is that these newly declassified experiments have not been in the public arena for a sufficient time to have been carefully and critically scrutinized. As with the original ganzfeld data base and the autoganzfeld experiments, it takes careful scrutiny and a period of a few years to find the problems of newly published or revealed parapsychological experiments. One obvious problem with the SAIC experiments is that the remote viewing results were all judged by one person -- the director of the program. I believe that Utts agrees with me that we have to withhold judgments on these experiments until it can be shown that independent judges can produce the same results. Beyond this, we would require, as with any other set of newly designed experiments, replication by independent laboratories before we decide that the reported outcomes can be trusted.

6. Prima Facie Evidence

Utts and other parapsychologists also talk about prima facie evidence in connection with the operational stories of the psychics (or remote viewers) employed by the government. Everyone agrees there is no way to evaluate the accounts of these attempts to use input from remote viewers in intelligence activities. This is because the data were collected in haphazard and nonsystematic ways. No consistent records are available; no attempt was made to interrogate the viewers in nonsuggestive ways; no contemporary systematic attempts to evaluate the results are there, etc.

The attempts to evaluate these operational uses after the fact are included in the American Institutes for Research (A.I.R.) report and they do not justify concluding anything about the effectiveness or reality of remote viewing. Some stories, especially those involving cases that occurred long ago and/or that are beyond actual verification, have been put forth as

evidence of apparently striking hits. The claim is that these remote viewers are right on -- are actually getting true psychic signals -- about 20 percent of the time.

Call it *prima facie* or whatever, none of this should be considered as evidence for anything. In situations where we do have some control comparisons, we find the same degree of hitting for wrong targets (when the judge does not realize it is the wrong target) as for the correct targets. A sobering example of this with respect to remote viewing can be found in David Marks and Richard Kammann's book *The Psychology of the Psychic* (Prometheus Books, Amherst, New York, 1980).

Psychologists, such as myself, who study subjective validation find nothing striking or surprising in the reported matching of reports against targets in the Stargate data. The overwhelming amount of data generated by the viewers is vague, general, and way off target. The few apparent hits are just what we would expect if nothing other than reasonable guessing and subjective validation are operating.

7. Consistency Among the Different Sources

Utts points to consistencies in effect sizes across the studies. More important, she points out several patterns such as bigger effect sizes with experienced subjects, etc. I do not have time or space to detail all the problems with these apparent consistencies. Many of them happen to relate to the fact that the average effect sizes in these cases are arbitrary combinations of heterogeneous sources. Moreover, where Utts detects consistencies, I find inconsistencies. I have documented some of these elsewhere; I will do so again in the near future.

Conclusions

When we examine the basis of Utts's strong claim for the existence of psi, we find that it relies on a handful of experiments that have been shown to have serious weaknesses after undergoing careful scrutiny, and another handful of experiments that have yet to undergo scrutiny or be successfully replicated. What seems clear is that the scientific community is not going to abandon its fundamental ideas about causality, time, and other principles on the basis of a handful of experiments whose findings have yet to be shown to be replicable and lawful.

Utts does assert that the findings from parapsychological experiments can be replicated with well-controlled experiments given adequate resources. But this is a hope or promise. Before we abandon relativity and quantum mechanics in their current formulations, we will require more than a promissory note. We will want, as is the case in other areas of science, solid evidence that these findings can, indeed, be produced under specified conditions.

Again, I do not have time to develop another part of this story. Because even if Utts and her colleagues are correct and we were to find that we could reproduce the findings under specified conditions, this would still be a far cry from concluding that psychic functioning has been demonstrated. This is because the current claim is based entirely upon a negative outcome -- the sole basis for arguing for ESP is that extra-chance results can be obtained that apparently cannot be explained by normal means. But an infinite variety of normal possibilities exist and it is not clear than one can control for all of them in a single experiment. You need a positive theory to guide you as to what needs to be controlled, and what can be ignored. Parapsychologists have not come close to this as yet.